**Transformers to Learn Hierarchical Contexts in Multiparty Dialogue for Span-based Question Answering(ACL 2020)**

기존 pre-train 모델들 => 대화를 잘 표현하기 어렵다 & 주어진 대화에 대한 QA Task에서 저조한 성능

1. 대화는 구어체인 경우가 많다.
2. 한 명의 저자가 관련이 있는 토픽에 관해쓴 wiki나 뉴스 기사와 달리 대화는 여러명의 발화로 구성

따라서 단순히 합치는 방법보다는 내부적으로 연결되어 있는 방식의 표현법이 필요

=> 컨텍스트를 더 잘 이해할 수 있는 토큰과 발화의 계층적 임베딩을 학습하는 트랜스포머 제안

**Main Idea**

새로운 pre-training task(토큰 단위, 발화 단위의 LM, 발화 순서 예측) => 임베딩 질을 대화체에 맞게 하여 높임

Pre-training으로부터 생성된 계층적인 임베딩의 이점 잘 이용 & span-based QA에 적합한 새로운 Multitask learning Fine-tuning 접근법 제안

BERT와 RoBERTa 능가